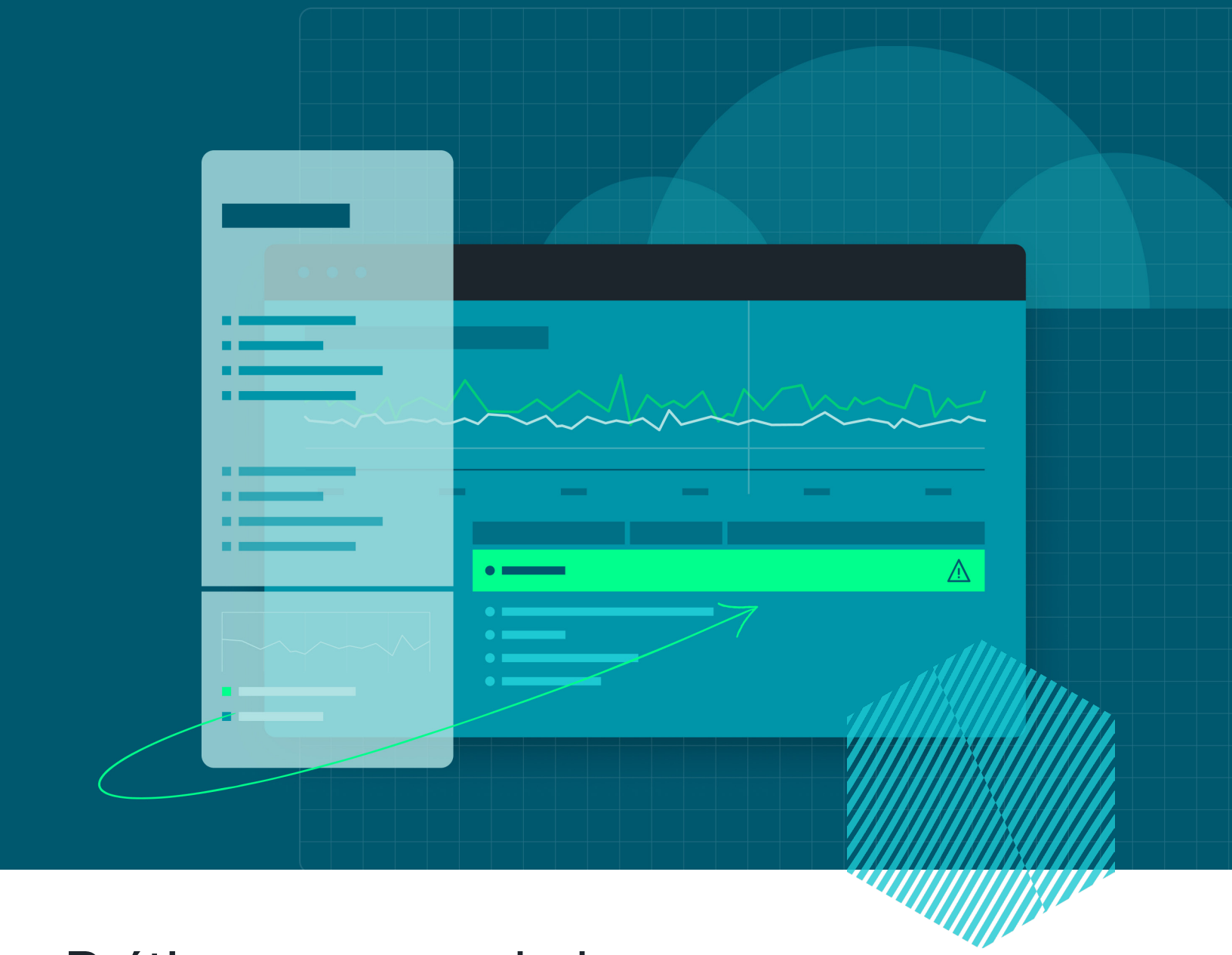
new relic



Práticas recomendadas de gerenciamento de logs

Tenha observabilidade full-stack com registros em log eficazes

White paper / Agosto de 2022

Conteúdo

Introdução

O gerenciamento de logs evoluiu. As organizações não precisam mais vasculhar montanhas de registros brutos de aplicativos e infraestrutura toda vez que algo para de funcionar. O gerenciamento de logs agora tem uma função central nas operações de negócios, inteligência e marketing das organizações. Logs fomentam a observabilidade. Logs bem estruturados ajudam organizações a entender rápida e facilmente como o sistema inteiro opera e, assim, prevenir problemas.

Mas o uso de logs para garantir observabilidade eficaz requer estratégia e cuidado. Não basta simplesmente despejar grandes quantidades de logs mal formatados em um banco de dados ou arquivo. Em vista disso, como as organizações podem mudar suas práticas de registro em log para melhorar sua capacidade de, com logs detalhados, correlacionar incidentes em diferentes aplicativos e infraestruturas, em tempo real, sem ter que alternar entre aplicativos e ferramentas? Como podem alcançar observabilidade total? Como podem chegar perto de atingir observabilidade full-stack para atender aos seus negócios?

É fácil mudar as práticas de registro em log para garantir que os logs aprimorem a observabilidade full-stack. Neste white paper, abordaremos algumas práticas recomendadas de registro em log para organizações modernas.

Registro em log tradicional

Tradicionalmente, o registro em log é feito em um silo de dados separado de outros sistemas. No passado, a observabilidade dependia de monitoramento do desempenho de aplicativos (APM) e de monitoramento de infraestrutura. Embora o monitoramento seja importante, ele não traz todas as informações que encontramos nos logs de aplicativos e dispositivos de infraestrutura. Muitas ferramentas de monitoramento e registro em log separadas e isoladas são focadas no aplicativo, que é apenas uma parte do stack, e não oferecem informações completas sobre o que e por que está acontecendo. É crucial que as equipes tenham as informações necessárias para acelerar o tempo de comercialização, obter insights melhores sobre o comportamento do cliente e reduzir o tempo de resposta a incidentes.

Muitas organizações escolhem não ter detalhes granulares de seus logs e sofrem para definir as causas subjacentes de seus problemas, ou usam ferramentas separadas para tentar mapear detalhes de log até erros e traces. Manter logs detalhados em silos separados impede as equipes de ter uma visão geral, aumenta os custos e o tempo de comercialização para produtos, reduz a visibilidade sobre a experiência do cliente e aumenta o tempo médio de resolução (MTTR).

Full-Stack Observability (Observabilidade Full-Stack)

A capacidade de observar tudo do stack de tecnologia que pode afetar a experiência do cliente é chamada de observabilidade full-stack ou observabilidade total. Ela se baseia em uma visualização completa de todos os dados de telemetria (métricas, eventos, logs e traces).

A observabilidade full-stack oferece visibilidade completa do desempenho de aplicativos e sistemas complexos, idealmente, a partir de uma única solução integrada, para resolver incidentes, reduzir o MTTR e entender a experiência do cliente.

Com a observabilidade full-stack, engenheiros e desenvolvedores não precisam lidar com amostras de dados, comprometer a visibilidade do stack de tecnologia ou perder tempo condensando dados dispersos. Assim, eles podem se dedicar ao desenvolvimento dos códigos mais importantes para a sua empresa.

Registro em log para ter observabilidade full-stack

Gerar logs para o stack inteiro pode ser difícil. Desenvolvedores e engenheiros podem ter dúvidas sobre o que registrar em log, quais detalhes incluir, e se dados demais podem gerar custos altos. Muitas organizações pagam caro para centralizar o seu gerenciamento de logs em uma plataforma diferente e acabam sendo forçados a limitar os dados de log enviados com base no desempenho e no preço. Essa amostragem de dados limita a visibilidade e o valor do negócio. Com isso em mente, apresentamos aqui algumas práticas recomendadas de gerenciamento de logs para observabilidade full-stack.

Registrar os logs certos

Logs são gerados pela gravação de textos em uma saída padrão ou um arquivo. A parte mais importante é decidir o que é registrado nesses logs. Logs devem incluir todos os metadados necessários para ajudar a rastrear eventos e causas raiz em uma investigação. Elementos de metadados de log podem incluir mensagens de erro ou stack traces e valores, métricas ou eventos relacionados.

Tudo que uma organização registrar em log deve ter um propósito. Sejam dados de uso, eventos de usuário ou erros de aplicativo e exceções, tudo que é registrado precisa ter valor para a equipe. Informações de dados de log devem:

- Ter uso imediato de alguma maneira
- Oferecer os detalhes necessários para entender as causas subjacentes e ajudar na tomada de decisões

Antecipar situações comuns

Logs não servem apenas para resposta a incidentes. Logs podem ajudar com outras áreas do negócio, como criação de perfis de desempenho ou coleta de estatísticas.

Fazer o registro de logs pensando em situações comuns garante que os logs ofereçam valor direto para a organização. Por exemplo, logs de interação de usuário podem oferecer insights cruciais sobre a experiência do cliente. Logs de sistema podem monitorar problemas ou falhas de hardware. Logs de aplicativo detalhados podem oferecer insights sobre desempenho e problemas potenciais como vazamentos de memória. Tudo isso pode ser importante na hora de tomar decisões de negócios.

Registrar mensagens significativas

Mensagens de log são tão valiosas quanto as informações e o contexto que oferecem. A inclusão de detalhes e contexto suficientes permite que as equipes usem os logs com eficácia. Infraestruturas de terceiros tendem a coletar os detalhes granulares necessários. No entanto, para aplicativos desenvolvidos internamente, as equipes devem coletar os detalhes de log que vão ajudá-las a diagnosticar e definir por que um erro/evento aconteceu para que possam tomar as medidas necessárias que afetam o negócio.

Para erros de aplicativos, a mensagem deve informar o que está acontecendo com aquela linha de código. Por exemplo, uma mensagem de erro `Falha na transação` não é tão útil quanto uma mensagem de erro `Falha na transação: não foi possível criar usuário ${path/to/file:line-number}`. Logs que incluem dados sobre transações ajudam desenvolvedores e engenheiros a entender porque as transações falharam.

Normalmente, códigos de erro ou status podem indicar o tipo de problema do aplicativo. Em vez de inserir apenas o texto do código ou número do erro, incluir uma descrição curta no log pode ajudar desenvolvedores e engenheiros na hora de pesquisar a resolução de problemas.

Logs devem oferecer informações essenciais para a organização. Desenvolvedores e engenheiros evitam registrar em log mensagens crípticas ou não descritivas que apenas alguns membros da equipe possam entender.

Manter logs simples e concisos

Embora seja essencial incluir informações suficientes na mensagem de log, o oposto também se aplica. Dados excessivos e desnecessários em uma mensagem podem inchar o tamanho e os custos de armazenamento de logs, desacelerar a busca de logs e desviar a atenção do problema central, dificultando a resolução de bugs.

Equipes devem manter logs concisos para coletar apenas as informações essenciais. Logs devem incluir a causa de um erro e excluir ruídos desnecessários.

Devem informar a causa raiz de um erro sem incluir todos os mínimos detalhes sobre o ambiente. Por exemplo, se um aplicativo falhou ao conectar e recuperar dados de uma API interna, pode ser útil registrar em log as mensagens de erro da API ou as informações do estado da rede do ambiente. Provavelmente será desnecessário incluir quanta memória o aplicativo usa ou quantos aplicativos estão sendo executados.

Não esquecer o timestamp

Equipes devem incluir um timestamp para os logs. Embora pareça óbvio, desenvolvedores e engenheiros que estão habituados a registrar logs em bancos de dados que incluem automaticamente data e hora podem não se lembrar de incluir um timestamp nas mensagens de log. Eles devem selecionar o nível granular que mais se adeque à situação e registrar nos logs. Tarefas de alta frequência podem exigir o registro de milissegundos, enquanto tarefas de baixa frequência podem rastrear apenas os minutos ou mesmo o dia. Mais importante que a granularidade é aplicar um padrão consistente por toda a organização.

Outra observação óbvia porém vital é sincronizar todos os sistemas ao mesmo tempo para que uma plataforma de observabilidade possa usar o timestamp para correlacionar eventos de log com outros dados de telemetria.

[1] (New Relic, Inc., n.d.)

Formatar logs de maneira analisável

Uma plataforma de observabilidade que não consegue extrair dados de logs não é muito útil. Equipes devem usar um formato de log que desenvolvedores e engenheiros possam analisar e manter uma estrutura de log consistente para facilitar a coleta e agregação. Por exemplo, o gerenciamento de logs da New Relic facilita na definição de regras de análise de log personalizadas,[1] mas regras de análise não servem de nada se os dados de log não forem inteligíveis.

Um bom exemplo de formato de log não analisável é o log de acesso padrão NGINX que contém texto não estruturado. É útil para pesquisa e só. Em um formato não analisável, equipes precisariam fazer uma busca de texto completa para responder à maioria das perguntas. Este é um exemplo de uma linha típica:

```
127.180.71.3 - - [10/maio/2022:08:05:32 +0000]
"GET /downloads/product_1 HTTP/1.1" 304 0 "-"
"Debian APT-HTTP/1.3 (0.8.16~exp12ubuntu10.21)"
```

Após a análise, o log é organizado em atributos, como `código de resposta` e `URL da solicitação`. Aqui está um exemplo da mesma informação de log em um formato analisável:

```
{
  "remote_addr":"93.180.71.3",
  "time":"1586514731",
  "method":"GET",
  "path":"/downloads/product_1",
  "version":"HTTP/1.1",
  "response":"304",
  "bytesSent": 0,
  "user_agent": "Debian APT-HTTP/1.3
  (0.8.16~exp12ubuntu10.21)"
}
```

Se o formato for totalmente personalizado, configurar o tipo de log ativará regras de análise definidas pelo cliente.

Se uma organização tem muitos aplicativos com a mesma função, as equipes devem focar em padronizar um formato de log para todos os apps. Isso facilita na incorporação de dados à plataforma de observabilidade, mesmo que a equipe associada a cada app queira ter visibilidade em diferentes atributos.

Uma análise detalhada de formatos de log

Há três categorias de formatos consistentes para estruturar um texto com implicações na usabilidade para a coleta de dados de uma ferramenta de agregação de logs. As três categorias de formatos são:

- **Estruturado**: um dos formatos estruturados mais comuns para o registro em logs é o JSON. Muitas ferramentas conseguem analisá-lo rapidamente. É muito flexível e leve. Idealmente, todos os logs seriam gerados em um formato estruturado. Embora o JSON ajude a organizar dados hierárquicos, outros exemplos de dados de log estruturados incluem formatos comuns, como valores separados por vírgulas (CSV) e valores separados por tabulação (TSV).
- **Comum**: um formato comum não é estruturado mas bem conhecido, definido e consistente. O formato de log comum da Apache para logs de acesso é um exemplo. A vantagem de um formato comum é que muitas ferramentas podem analisar dados prontos para uso.
- **Personalizado**: se um aplicativo não registra logs no formato estruturado ou comum, ele o faz no formato personalizado. Para reconhecer o início e o fim de uma linha de log individual durante o encaminhamento de logs, equipes podem ter que analisar. Criar regras de análise definidas pelo cliente pode ajudar a tornar os dados mais valiosos.

Categorizar e agrupar logs

Especificar um modelo de dados para logs ajuda equipes a fazer buscas de maneira mais eficaz. Elas devem definir e incluir atributos quando possível para categorizar e agrupar logs corretamente.

Os padrões de OpenTelemetry para logs criados por uma coalizão de líderes do setor, incluindo a New Relic, abordam muitos elementos como convenções de nomeação e definições de valores de campo.[2] Embora nem todo framework ofereça suporte nativo a logs formatados exatamente nesses padrões, eles podem servir como guia.

Atributos comuns que podem ser úteis em um modelo de dados de log incluem recursos, logs contextualizados e níveis de log.

Recursos
Recursos definem quando e de onde vieram os logs, tais como:

- Data e hora
- Nome do host da máquina ou identificador
- Nome do aplicativo ou serviço

O nome do host pode ser significativo em logs de aplicativos baseados em host clássicos com ambientes nomeados. Uma ID de pod ou contêiner organizaria melhor os logs de ambientes orquestrados ou conteinerizados.

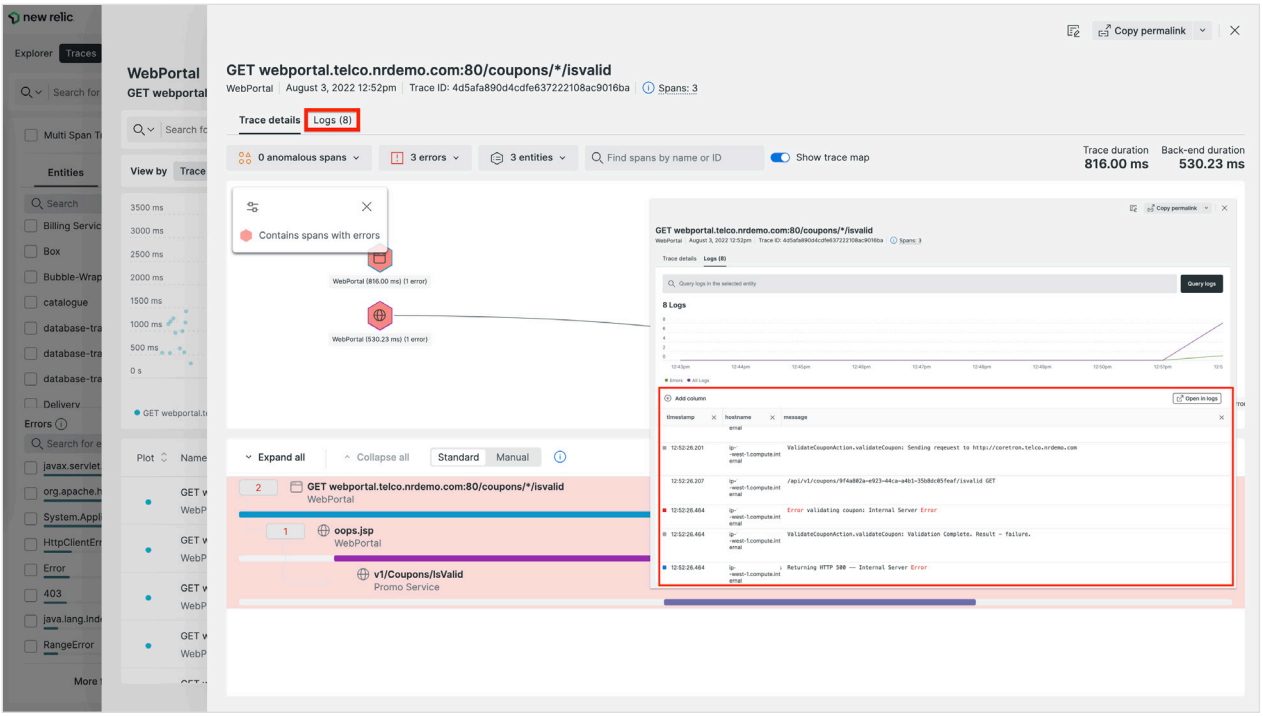
Ambientes orquestrados ou Platform-as-a-Service (PaaS) costumam popular logs com muitos metadados de maneira automática. Isso é ótimo para organizações, mas também é importante anotar os logs com qualificadores úteis que um sistema não tem como saber, como números de versão de produto, ambientes de preparação versus de produção, branches de teste ou versões de teste A/B. Agregação de logs significa que todos os logs de múltiplas fontes são coletados no mesmo sistema. Sem os metadados certos, equipes não conseguem distinguir um log de erro real em produção de uma transação que falhou como parte de uma execução de teste.

Outro recurso que pode ajudar equipes a identificar a fonte do log é o encaminhamento de log. Por exemplo, a maioria das soluções de encaminhamento de log oferecidas pela New Relic inclui automaticamente nas anotações de dados o tipo e a versão da ferramenta usada para entregar os dados.

[2] (OpenTelemetry, n.d.)

Logs contextualizados

Ver logs contextualizados de problemas nos aplicativos e hosts é útil para as equipes. Por exemplo, o recurso logs contextualizados da New Relic pode incluir automaticamente informações de aplicativos em logs. O agente New Relic APM oferece gerenciamento de dados de desempenho do aplicativo para o framework de registro em logs e os inclui nos logs de aplicativo. Assim, os logs contextualizados correlacionam automaticamente dados de log com eventos e traces de aplicativo associados. Erros de APM e distributed traces vinculam-se diretamente aos logs criados durante a mesma transação do evento ou trace. Logs contextualizados criam essa correlação inserindo IDs de span, trace e nome do aplicativo nas mensagens de log. Assim, equipes podem unir dados de aplicativo e log e resolver problemas com muito mais rapidez.



Logs filtrados para mostrar erros contextualizados de trace na plataforma de observabilidade da New Relic

Níveis de log
Desenvolvedores, utilizadores DevOps e gerentes podem se referir a níveis de log como níveis de severidade. Níveis de log descrevem a importância relativa do evento (com termos como debug, info, warning, error e fatal) e o nível de densidade da informação a partir do framework de registro em log. Um atributo de severidade ajuda a filtrar ou descartar informações menos críticas para que os termos busquem apenas por erros críticos.

O uso eficaz de níveis de log pode limitar a quantidade de dados, reduzir o custo de usar uma ferramenta de gerenciamento de logs centralizado e manter a agilidade dos resultados de busca. Em alguns casos, é possível que não se consiga controlar como os aplicativos gerenciam logs, porém, o sistema de gerenciamento de logs pode idealmente descartar dados indesejados. Por exemplo, com a New Relic, equipes podem descobrir anomalias usando padrões orientados por machine learning com base no nível de log. Níveis de log classificados por cor também oferecem um indicador visual para direcionar a atenção para as áreas mais críticas.

Equipes devem usar os níveis de log com cuidado, principalmente o nível de log resolução de bugs. A resolução de bugs pode ajudar a coletar mensagens muito prolixas associadas a um comportamento específico, mas a resolução de bugs desnecessária pode criar um volume significativamente alto de logs e tornar as funções de ingestão e busca mais lentas sem agregar valor. Equipes e projetos maiores podem se beneficiar dos padrões de nível de log para métodos de agrupamento, categorização e gerenciamento de logs consistentes.

Usar ferramentas e framework de registro em log

Em vez de gastar tempo e recursos implementando uma solução de log do zero, usar frameworks e ferramentas de registro em log estabelecidas e bem testadas pode poupar tempo e dor de cabeça. Por exemplo, os agentes de linguagem New Relic APM decoram logs com os metadados necessários para oferecer acesso ao recurso automático logs-in-context (logs contextualizados) e encaminhar logs sem precisar instalar ou manter um software de terceiros, tudo em uma única implantação.

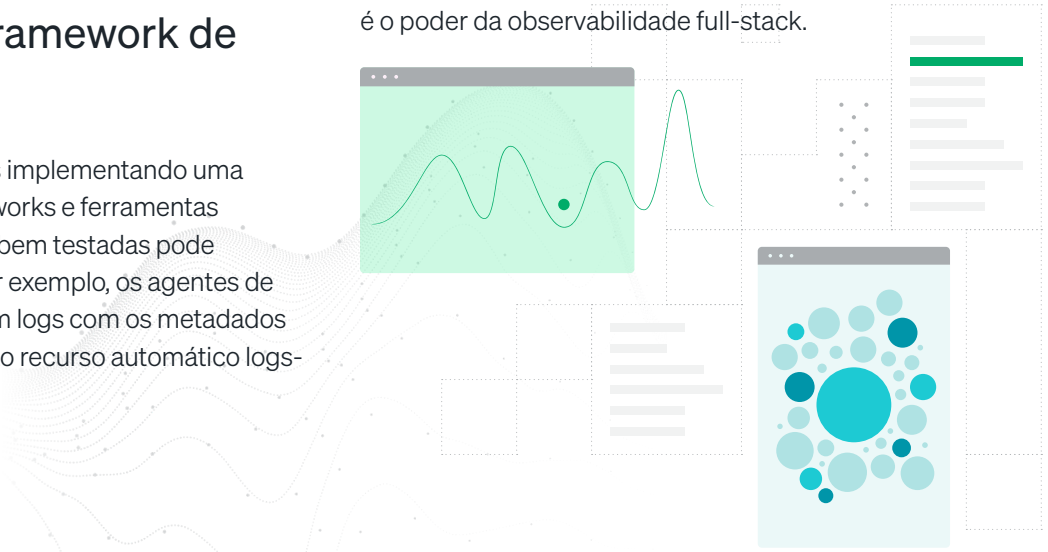
Usar um framework de registro em log consistente simplifica a adoção pela equipe de engenharia, normaliza a saída de log e garante que as equipes possam ativar os logs contextualizados de maneira uniforme. Equipes também devem ser cautelosas ao introduzir frameworks de registro em log e testar o impacto do desempenho assim como fariam com qualquer código novo.

Referenciar valores grandes sem incluí-los

Em alguns casos, equipes precisam de grandes blocos de dados do log para oferecer contexto aprofundado, como um registro de memória ou um conjunto de arquivos ou imagens. Em geral, é melhor salvar os dados separadamente ou até mesmo carregá-los em um servidor designado e referenciar a localização no log do que salvar como um borrão no log. Equipes devem manter logs bem leves e acessar os dados separadamente.

Compartilhar vistas, consultas e alertas úteis

Equipes devem criar e compartilhar exibições, consultas e alertas padronizados para os seus logs para oferecer insights mais completos sobre o estado atual da sua organização e aumentar a visibilidade e a comunicação entre equipes. Esse é o poder da observabilidade full-stack.



O que não registrar em log

É tentador registrar tudo que possa ser útil, mas as equipes devem evitar algumas exceções e armadilhas.

Informações sensíveis

Equipes devem lidar com informações sensíveis com cuidado. É essencial proteger dados regulados, como informações de identificação pessoal (PII) e números de cartão de crédito, de acordo com leis e regulamentos, como o Regulamento Geral sobre a Proteção de Dados (GDPR) na União Europeia[3] e a Lei de Responsabilidade e Portabilidade de Seguro Saúde (HIPAA) nos Estados Unidos.[4]

O guia de registro em log Open Web Application Security Project (OWASP) especifica o que não deve ser registrado em logs, como tokens de acesso, senhas, informações sensíveis e informações que indivíduos desejem resguardar.[5]

Logs armazenados em um servidor ou banco de dados privado podem facilmente registrar PII, como nomes e endereços de email, por acaso. Para rastrear ações ou eventos de um usuário específico, equipes devem usar identificadores anônimos. Embora dados de log estejam seguros em uma plataforma de observabilidade como a da New Relic, é importante tomar muito cuidado com a transmissão de PII para fora da organização.

Código-fonte e dados proprietários

Além de informações regulatórias e de compliance, equipes podem querer evitar armazenar informações sensíveis nos logs, como código-fonte de aplicativos ou dados protegidos dentro da organização.

Além de armazenar logs de maneira segura, também é importante assegurar o acesso a eles. Informações que podem revelar segredos comerciais ou projetos e recursos não anunciados ou não lançados não devem constar em logs. Portanto, equipes devem eliminar essas informações dos logs, principalmente se armazenarem logs externamente ou em um serviço de terceiros.

Informações duplicadas

Adicionar informações duplicadas a logs não gera problemas, e ter informações a mais costuma ser melhor do que não ter o suficiente. No entanto, incluir informações duplicadas pode criar logs desnecessários que não servem para nada, o que gera mais custos.

[3] (European Commission, n.d.)
[4] (U.S. Department of Health and Human Services (HHS), n.d.)
[5] (Open Web Application Security Project (OWASP), n.d.)

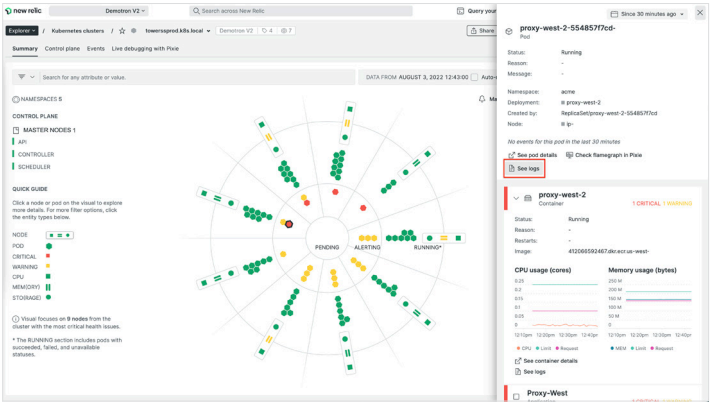
Conclusão

Registrar logs úteis para que melhorem a observabilidade full-stack ajuda a tomar decisões em tempo real que afetam o negócio e garantem que desenvolvedores e engenheiros passem menos tempo resolvendo bugs e respondendo a incidentes e mais tempo focando na inovação.

Com estas práticas recomendadas em vigor, logs podem oferecer os detalhes necessários para você manter tudo funcionando para os clientes, ter mais visibilidade do stack completo para resolver problemas mais rápido e acelerar o desenvolvimento.

Plataforma de observabilidade da New Relic

A New Relic oferece uma plataforma unificada para todos os dados de telemetria, incluindo logs detalhados. A plataforma de observabilidade da New Relic incorpora gerenciamento de logs, APM, monitoramento de infraestrutura, monitoramento Serverless, monitoramento de Mobile, monitoramento de Browser, monitoramento sintético, distributed tracing, monitoramento de Kubernetes, e mais. Esses recursos ajudam organizações a visualizar, analisar e resolver problemas de todo o stack de software. Como parte da plataforma, o gerenciamento de logs da New Relic ajuda organizações a combinar dados de log com dados de aplicativo e monitoramento de infraestrutura, o que resulta em uma plataforma de observabilidade completa e poderosa.



APM, infraestrutura, evento e acesso a logs combinados em uma única vista

A New Relic une métricas, eventos, logs e traces de todo o stack de software integrado com AIOps (artificial intelligence for IT operations), o que ajuda organizações a buscar logs mais rápido e é mais economicamente acessível que soluções legadas diferentes. Em vez de usar ferramentas separadas em diferentes partes do stack, desenvolvedores e engenheiros podem ver todos os logs detalhados relacionados a um erro específico com facilidade e em uma vista unificada.

Problemas de agilidade e escalabilidade em soluções de log legadas são um desafio para a consulta de logs detalhados, pois a busca com dados atrasados pode demorar minutos ou mesmo horas. Em contraste, a busca no gerenciamento de logs da New Relic leva apenas segundos, então a investigação e a resposta a incidentes por todo o stack de software é a mais rápida possível.

A plataforma de observabilidade da New Relic inclui gerenciamento de logs, um nível gratuito para clientes com baixo volume e um preço baixo por GB que permite a equipes ingerir todos os logs detalhados de que precisam.

Para começar a usar o gerenciamento de logs da New Relic, cadastre-se em uma conta gratuita hoje mesmo. Sua conta gratuita inclui ingestão de 100 GB de dados por mês, um usuário Full Platform e um número ilimitado de usuários básicos.

Cadastre-se agora

Referências

European Commission. n.d. “EU data protection rules.” European Commission. Accessed July 19, 2022. https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en.

New Relic, Inc. n.d. “Parsing log data.” New Relic Documentation. Accessed July 28, 2022. https://docs.newrelic.com/docs/logs/ui-data/parsing/#custom-parsing.

OpenTelemetry. n.d. “OpenTelemetry Logging Overview.” OpenTelemetry. Accessed July 18, 2022. https://opentelemetry.io/docs/reference/specification/logs/overview/.

Open Web Application Security Project (OWASP). n.d. “OWASP Logging Guide.” https://owasp.org/www-pdf-archive/OWASP_Logging_Guide.pdf.

U.S. Department of Health and Human Services (HHS). n.d. “Summary of the HIPAA Security Rule.” HHS.gov. Accessed July 19, 2022. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

new relic.